**Clonal Evolution of Paediatric Burkitt Lymphoma Through Time and Space**

Alexander S. Steemers1,2, Mark J. van Roosmalen1,2, Rico Hagelaar1,2, Laurianne Trabut1,2, Mark Verheul1,2, Friederike Meyer-Wentrup1, Ruben van Boxtel1,2

1. Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
2. Oncode Institute, Utrecht, The Netherlands

**Abstract**

**Introduction**

Cancer arises due to the stepwise accumulation of mutations in key driver genes. Such mutations typically accumulate over an extended period, commencing even decades before clinical presentation1,2 . **Paradoxically, however, cancers such as Burkitt lymphoma (BL) predominantly affect children**, with a median age of onset of 10 years3 **.** **This raises intriguing questions about how oncogenic processes unfold over such a compressed timescale.**

**Paediatric** BL (pBL) is a highly aggressive, germinal centre-derived B-cell malignancy and is regarded as the fastest-growing human tumour4 . A distinctive feature of pBL is its primary extranodal presentation, most commonly involving the ileocecal region. While localized disease often responds well to intensive chemotherapy, both acute and long-term side effects are common5,6 . In addition, involvement in the bone marrow or central nervous system (CNS) is stratified as advanced-stage BL, while recurrent/refractory disease is associated with an extremely poor prognosis7 . A deeper understanding of the molecular mechanisms driving lymphomagenesis, dissemination and relapse may pave the way for more targeted and less toxic therapeutic strategies.

The genetic hallmark of BL is a reciprocal translocation that places the *MYC* oncogene under the control of immunoglobulin enhancers, resulting in constitutive MYC overexpression8 . MYC dysregulation alone is insufficient for BL oncogenesis, and additional genetic alterations are required to fully drive malignant transformation9–11 . These cooperating mutations typically involve genes important in B-cell receptor (BCR) signalling (*ID3*, *FOXO1*, *TCF3*), cell survival and proliferation (*MYC*, *CCND3*, *TP53*), and in SWI–SNF chromatin remodelling (*SMARCA4*, *ARID1A*), highlighting recurrent disruption of signalling, growth regulation, and epigenetic networks in BL12 . Despite advances in defining the molecular drivers, the timing and clonal architecture of BL mutations remain largely unknown, mainly because conventional bulk sequencing blurs signals from genetically heterogeneous subclones.

Here, we analysed the genomes of malignant and non-malignant B-cells from a cohort of pBL patients using single-cell whole genome sequencing (scWGS). Since primary lymphoma cells do not clonally expand *ex vivo*, precluding the generation of single-cell colonies, we implemented the novel Primary Template-directed Amplification (PTA) technique which allows for direct whole genome amplification of individual cells13,14 . Tumour cells were sampled either from a single anatomical site or from multiple distinct locations within the same patient, allowing us to reconstruct both the temporal and spatial trajectory of pBL at single-cell resolution.

**Results**

Previously (from the Machado paper) it was thought that BL cells have the same mutational burden as normal memory B cells, is that true? 🡪 No, they have higher mutational burden than expected when you consider both shared and private somatic mutations of individual cells. Well, it's also not untrue. However, they compared the MRCA. To know the complete mutation burden, one needs to perform scWGS.

**Figure 1: Experimental overview and ageline to show increased mutational burden in BL cells.**

1. Overview of experimental pipeline for scWGS
2. Overview of patients and mutations
3. Ageline with myc-SBS-, myc-SBS+ and myc+ cells (call them Malignant cells, Normal Memory B cells, normal Naïve B cells in the actual figure) Add the MRCAs to show that Machado et al were not wrong. Also, you can use this later for the modeling when you generate ultrametric trees. You could add the entire cohort here. (bone marrow not in ageline model)
4. Ageline with Malignant cells and Bulk WGS? Per signature
5. Mutational signatures? Would be nice to give a general overview. Also, what is the cause of the increased mut burden in the single cells? The observation that the mut burden at the single cell clevel is higher than the MRCA is the main message of this figure. Also, this will also allow you to show that we don’t have an artifact signature left (which is important, because a reviewer could argue that this increase is caused by that.)
6. Oncoplot of the entire cohort including single cells? This will show that the patients you’ve chosen for scWGS are representative for the cohort. -> although increased mut burden in single-cells compared to MRCAs, we don’t see additional hits in the single cells. What does that mean?

Okay we see higher mutational burden, but what is the clonal architecture of these tumours? Which of these mutations are clonal and which are sub-clonal? Are there subclonal drivers? Can we time CNVs in respect to drivers? We will look into a larger cohort, in this case 15 additional bulk WGS samples

**Figure 2: Intra-tumoral heterogeneity. 🡪 main message order of drivers**

1. Absolute tree PIA9ileocecal region with drivers and CNAs
2. Absolute tree PVA9 ileocecal region with drivers and CNAs
3. Absolute tree PJBU ileocecal region with drivers and CNAs
4. Absolute tree P3G6 ileocecal region with drivers and CNAs
5. Absolute tree PRN4 lymph node with drivers and CNAs
6. Absolute tree P856 pleural cavity with drivers and CNAs
7. Mutational sigs early vs (intermediate vs) private- > message? The mut processes are different during early development? This can be coupled back to Fig. 1D
8. MutationalTimeR to time driver mutations compared to CNAs -> validate these findings in the entire cohort. It will be very interesting to do this in the tree and estimate in actual time when these events happened (in the next figure). (look at VAF 0.3, 06)

Can we now time the clonal expansion? Can the MRCA be detected years or months before sampling? It is expected to be recent considering the clinical manifestation, but does that still hold true? Why is this important? We observe that there are quite a few unique mutations per cell. Again, couple back to comparison mut burden MRCA and individual cells. Mutation accumulation is associated with time, what does this mean for the initiation of this disease? -> therefore, we build the ultrametric trees.

**Figure 3: Timing tumorigenesis in years.** -> What’s the message of this figure?

1. SBSblood or C>T to show that one constant mutation rate in normal and in malignant cells (could be supp)
2. Ultrametric tree 27774 with drivers and CNAs
3. Ultrametric tree 28306 with drivers and CNAs
4. Ultrametric tree 27548 with drivers and CNAs
5. Ultrametric tree P3G6 with drivers and CNAs
6. Ultrametric tree PB08410 with drivers and CNAs
7. Ultrametric tree PB14458 with drivers and CNAs

So far we have shown the clonal evolution within single sites, particularly the abdomen, which is the primary site of BL. Bone marrow involvement commonly occurs in sporadic BL and includes leukaemia with extensive blasts (>25%) in the marrow. Bone marrow involvement is present in fewer than 10% of patients at initial diagnosis but frequently occurs as a recurrent or treatment-resistant disease complication. Indeed, bone marrow involvement is associated with worse outcomes. To understand inter-tumoral heterogeneity and migration we studied two patients, one with bone marrow dissemination at diagnosis and one patient with bone marrow involvement at relapse. Can we now answer the question, which clone is responsible for the migration, and when do the two clones diverge? Also, what we learned with sampling multiple sites is that the myc:igh translocation is NOT the limiting step for BL, since distinct sites acquire driver mutations independently. We can also time the clonal expansions. We learned a bit more about what is the rate limiting step.

**Figure 4: Inter-tumoral heterogeneity. (maybe fig 5)**

1. Absolute tree PB08410 LN + BM with drivers and CNAs
2. Absolute tree PB14458 PL + BM with drivers and CNAs
3. MIXCR to show clones are same/different (Supp)
4. Mutational signatures between sites? If interesting
5. Ultrametric PB08410
6. Ultrametric PB14458
7. Ultra deep sequencing result at diagnosis BM?

Finally, since we have data paediatric patients with a range of ages (4Y, 6Y, 12Y, 13Y, 14Y, 17Y), can we check whether age influences things like growth rates/doubling times and latency periods. This probably also depends on our timing of the first hits using a combination of treebuilding and mutationaltimer. If we can indeed show that this is in all patients around the same age, it would be interesting to see what differs. Is there a correlation? Based on previous findings (https://www.nature.com/articles/s41586-025-08817-2), we expect a trend of younger age of onset correlating with more explosive growth together with a shorter duration between the beginning of clonal expansion and diagnosis. Is this what we see in BL as well?

**Figure 5: Growth rates and correlation with age. (maybe fig 4)**

1. Growth rates vs latency period
2. Doubling time vs latency period

**Discussion**

**Methods**

**Patient samples**

Samples were collected from patients with Burkitt lymphoma via the biobank of the Princess Máxima Center for Pediatric Oncology in accordance with the Declaration of Helsinki. Patients provided informed, written consent for the use of their samples for research. In addition, bulk WGS data (tumour and matched normal samples) from 15 paediatric BL patients was obtained from our in-house diagnostics database, where samples are routinely sequenced as part of clinical evaluation. The study was covered under the proposal PMCLAB2022-303 approved by the Institutional Review Board of the Máxima. Clinical information of the patients, including which sites where sampled, is summarized in Supplementary Tables X.

**Sample work-up**

Samples were worked up for single-cell WGS according to our previously published STAR Protocol guidelines15 . Briefly, samples were stained for fluorescence-activated cell sorting (FACS) after thawing. Burkitt lymphoma (BL) cells and normal B-cells were purified on a SH800S Cell Sorter (Sony) using a 100 μm microfluidic sorting chip. First, manual gates were used to remove debris (FSC-A vs. SSC-A), doublets (FSC-A vs. FSC-H) and dead cells (DAPI+). BL cells were identified using the following surface markers: CD20+CD10+ and either IgK+ or IgL+ (depending on their monoclonal immunoglobulin light chain rearrangement). Normal B cells were identified using the following surface markers: CD20+CD10-. Single BL cells and normal B cells were index sorted in a 96-well plate or 384-well plate prepared with PTA-buffer. Bulk BL cells were then sorted for DNA isolation. Representative gating strategies are shown in Supplementary Fig. X. All antibodies were obtained from BioLegend. Antibodies used for BL and normal B-cell populations: CD20-FITC (clone 2H7, 1:25, #302303), CD10-APC (clone HI10a, 1:50, #312209), IgK-PE/Cy7 (clone MHK-49, 1:50, #316520), IgL-AF700 (clone MHL-38, 1:25, #316631).

**Germline controls**

Mesenchymal stromal cells (MSCs) were cultured from the bone marrow fraction of 500,00 cells/well in 12-well culture dishes with 2 mL Advanced DMEM-F12 medium (#12634010, Gibco) supplemented with 10% FBS (Gibco), 1% GlutaMax (#35050061, Gibco) and 1% Penicillin-Streptomycin (#15140122, Thermo Fisher Scientific). Medium was refreshed every other day to remove non-adherent cells and MSCs were be harvested when confluent (after approximately 3 weeks).

**DNA isolation and WGS**

DNA was isolated from cell pellets of bulk BL cells and MSCs using the DNeasy DNA Micro Kit (#56304, Qiagen), following the manufacturer’s instructions. The standard protocol was slightly adjusted by adding 2 µL RNase A (#19101, Qiagen) during the lysis step and eluting DNA in 50 µL low EDTA TE buffer (10 mM Tris, 0.1 mM EDTA, G Biosciences, #786150).

DNA from single cells was amplified using the ResolveDNA® WholeGenome Amplification Kit (#100136, BioSkryb) or the ResolveDNA® Whole Genome Amplification Kit v.2.0 (#100545, BioSkryb) using a D100 Single Cell Dispenser (HP), according to according to the manufacturer’s instructions. Details on which kit was used per sample is summarized in Supplementary Table X.

For each sample, DNA libraries for Illumina sequencing were generated from at least 45 ng genomic DNA using standard protocols. For PTA-amplified DNA at least 200 ng genomic DNA was used. The libraries were sequenced at a depth of 15x (single cells) or 30x (bulk tumour and MSC samples).

**Read mapping**

Sequencing reads were first mapped to genome GRCh38 using the BWA-MEM2 (Burrows–Wheeler Aligner) algorithm (v.2.2.1) using the settings “–M –c 100 -R”. Duplicates were then marked using GATK4spark (v.4.4.0.0), and base recalibration was performed using GATK4 (v.4.4.0.0).

**Mutation calling, filtering and annotation**

Mutation calling was performed using GATK’s HaplotypeCaller on combined samples per patient. Variant quality score recalibration (VQSR) was applied separately for SNPs and INDELs using the VariantRecalibrator tool with the following annotations: -an QD -an MQ -an MQRankSum -an ReadPosRankSum -an FS -an SOR -mode SNP for SNPs, and -an QD -an MQ -an MQRankSum -an ReadPosRankSum -an FS -an SOR -mode INDEL for INDELs. ApplyVQSR was then run with --truth-sensitivity-filter-level 99.9 -mode SNP and --truth-sensitivity-filter-level 99.9 -mode INDEL.

Next, GATK’s VariantFiltration was run with the following options: “--filter-expression 'QD < 2.0' --filter-expression 'MQ < 40.0' --filter-expression 'FS > 60.0' --filter-expression 'HaplotypeScore > 13.0' --filter-expression 'MQRankSum < -12.5' --filter-expression 'ReadPosRankSum < -8.0' --filter-expression 'MQ0 >= 4 && ((MQ0 / (1.0 \* DP)) > 0.1)' --filter-expression 'DP < 5' --filter-expression 'QUAL < 30' --filter-expression 'QUAL >= 30.0 && QUAL < 50.0' --filter-expression 'SOR > 4.0' --filter-name 'SNP\_LowQualityDepth' --filter-name 'SNP\_MappingQuality' --filter-name 'SNP\_StrandBias' --filter-name 'SNP\_HaplotypeScoreHigh' --filter-name 'SNP\_MQRankSumLow' --filter-name 'SNP\_ReadPosRankSumLow' --filter-name 'SNP\_HardToValidate' --filter-name 'SNP\_LowCoverage' --filter-name 'SNP\_VeryLowQual' --filter-name 'SNP\_LowQual' --filter-name 'SNP\_SOR' -cluster 3 -window 10”.

Annotation of variants was done using Ensembl Variant Effect Predictor (VEP) (v.112.0) with settings = "--vcf --plugin AlphaMissense,file --plugin NMD --plugin dbNSFP". A full pipeline description is available at https://github.com/ToolsVanBox/ASAP. The version of ASAP used was v.1.0.4.

**Somatic mutation filtering**

Somatic mutation calls were generated using SMuRF (v3.0.0) to ensure high-confidence variant identification (available at www.github.com/ToolsVanBox/SmuRF). Variants were retained if they met the following criteria: (A) assigned a GATK Phred-scaled quality score of at least 100, (B) had a mapping quality of ≥60 for 30x coverage data or ≥55 for 15x coverage, (C) had minimum base coverage of 10 (30x) or 5 (15x), (D) exhibited a GATK genotype quality of 99 (for heterozygous calls) or 10 (for homozygous calls) in both the tumour sample and matched normal, and (E) showed no supporting reads in the matched control. Finally, mutations with low variant allele frequencies (VAF) were removed to obtain a set of clonal mutations. The VAF cut-off for 15x sequenced single cells was 0.15, as mutations below this cut-off could be technical artifacts. The cut-off for 30x sequenced bulk tumour samples with no contamination of healthy cells was 0.3, as mutations below this cut-off are subclonal. For 30x sequenced bulk tumour sample PRN4GBDLBC72 with ~85% purity the cut-off was 0.25. The cut-off for 90x diagnostic bulk samples was chosen based on tumour purity; 0.25 for samples with >70% purity and 0.20 for samples with 60%-75% purity. Again, mutations below these cut-offs are likely subclonal or technical artifacts. VAF distributions of the somatic mutations of all samples are shown in Supplementary Fig. X.

**Mutation filtering for PTA single-cell WGS data**

For single cells, our in-house developed pipeline PTATO (v.1.3.3) was applied, which takes the VCF files produced by SMuRF (v.3.0.0) and filters these using germline mutations combined with a random forest and walker to separate real somatic mutations from amplification-induced artifacts. For a full description of this pipeline, see our previously published publication16 .

**Quality control of single-cell WGS data**

To ensure high-confidence single-cell genomic profiles, we first excluded samples with a callable loci fraction below 0.45 (n=13). The cutoff was determined based on the distribution of the callable loci fraction among all samples (Supplementary Fig. X). Next, for the remaining samples, we fitted a logistic regression model to describe the relationship between mean coverage and callable genome fraction and flagged samples with residuals below −0.05 as having low mapping quality (n=37) (Supplementary Fig. X). Again, the cutoff here was selected based on the distribution (Supplementaty Fig. X). Finally, with the remaining samples we assessed variant-level quality by calculating the total variation distance (TVD) between each sample’s VAF histogram and the cohort median distribution. Samples with a TVD exceeding the median + 3 × MAD were flagged as outliers (n=9) (Supplementary Fig. X). Samples failing any of these criteria were excluded from downstream analyses (total n=59).

**Per sample MYC::IGH mutation status**

Patient-specific primers were designed that flank the translocation breakpoints to assess the presence or absence of the MYC–IGH rearrangement (Supplementary Fig. X). In addition, the MYC–IGH translocation was manually reviewed and validated using IGV (v.2.8.2)18 . Samples were classified as translocation-positive if at least one pair of reads spanned the breakpoint region (Supplementary Fig. X).

**B-cell receptor repertoire**

BAM files for each single-cell or bulk sample were first down-sampled to the TCR and BCR loci using a custom BED file. The locus-specific BAMs were then converted to paired-end FASTQ files. Receptor reconstruction was performed with MiXCR (v.4.3.2)19 . For each clone, the top-scoring V, D and J genes, as well as the amino acid CDR3 sequence, were extracted. Clones with CDR3 sequences containing incomplete (“\_”) and/or stop (“\*”) codons were filtered out from downstream analysis. BCR repertoires were visualized using the R package ComplexHeatmap v.2.22.020 .

**Phylogenetic tree construction**

Phylogenetic reconstruction of malignant and non-malignant single cells from the same patient was performed using CellPhy (v.0.9.254)21 , which relies on RAxML-NG for maximum likelihood inference. CellPhy considers allelic dropout and amplification errors on a per-sample basis to estimate the most likely tree structure. Analyses were performed using phred-scaled genotype likelihoods (PL) under the default “GT16 + FO + E” model. For each tree, 100 bootstrap iterations were generated, and the proportion of replicates supporting each split is shown in Figs. X and X, providing a measure of node confidence.

Trees were rooted using the “root” function from the treeio R package, with MSCs used as the outgroup. To validate tree topology, variant allele frequencies (VAFs) from bulk tumour samples were assessed. Mutations on clonal branches showed VAFs around 0.5, consistent with their presence in all tumour cells, while most private mutations were undetectable in bulk, indicating they were restricted to single malignant cells.

To account for differences in sequencing sensitivity between cells, we corrected branch lengths using callable loci information. Specifically, we calculated a node-specific detection power, , as , where is the callable-loci fraction of sample , and scaled the original CellPhy branch lengths by this value to yield callable-normalised branch lengths.

**Mutational burden**

* Used only somatic mutations in autosomal chromosomes from **CellPhy output!**
* Used linear model to find line (slope + y-intercept)

**Driver annotation**

Single base substitution and indel driver events were extracted from the HaplotypeCaller output. Only mutations with VEP-assigned impact of MODERATE, HIGH, or Likely pathogenic were considered (Supplementary Table X). To focus on bona-fide Burkitt-lymphoma drivers, the variant list was filtered against a curated set of 81 known Burkitt lymphoma driver genes, compiled from the IntOGen database and a large previously published WGS study of BL by Panea et al22 (Supplementary Table X). Manual inspection in IGV (v.2.8.2) was performed to validate true somatic drivers and rescue any potential false negatives. The final, curated mutation matrices were visualized using oncoplots generated with the ComplexHeatmap package in R (v.2.22.0)18,20 (Supplementary Fig. X). Driver events (including SBS, Indels, SVs and CNVs) were manually added to the phylogenetic tree branches. If a driver mutation was not detected in a tumour cell(s), we followed Occam’s razor and attributed the absence to technical limitations - low coverage or allelic dropout – choosing the most parsimonious explanation based on the phylogenetic tree reconstruction. In the phylogenetic trees, only driver mutations which were absent from the patient-matched normal B-cells were included and highlighted in colour.

**Mutational signature extraction and refitting**

* First performed De novo signature extraction with MutationalPatterns (using reference datasets in addition to ours...Machado, bulk Lymphomas) using **CellPhy output VCF files!**
* Then checked to which known signatures these are like
* Then did strict refit with bootstrap on a per-sample basis using the de novo signatures
* Check for cosine similarity original vs reconstructed and made sure all are above 0.85
* Mutational signatures extraction and refitting of phylogenetic trees

**Copy number aberrations**

All copy-number analyses were performed in R (v.4.3.1). For each sample, 1 Mb binned read counts were normalised to the sample mean and scaled to the assumed ploidy (default = 2). To smooth local fluctuations, we applied a centred rolling window and retained the window median, start, end and mid-point coordinates for each chromosome using the Roll package (v.1.1.7). Cytoband coordinates (hg38) were intersected with the smoothed data to assign each bin to the p- or q-arm. Per chromosome-arm, copy-number ratios were clustered with a Gaussian mixture model using mclust (v.6.1.1); the number of components was set to the number of modes identified by LaplacesDemon (v.16.1.6). Component means were rounded to the nearest integer to define discrete copy-number states, and each bin was re-labelled with its most probable state. Bins with corrected ratios ≥ default ploidy + 0.8 were called gains while those with corrected ratios ≤ default ploidy – 0.8 were labelled as losses. Consecutive bins sharing the same state were merged into segments, producing a wide sample × bin matrix. Heatmaps were generated with ggplot2 (v.3.5.2) and faceted by chromosome + arm using ggh4x (v.0.3.0) with a capped colour gradient (0–4 copies).

**Timing branches**

* Rtreefit package using the Ultrametdcc
* Where I put the switch
* Early growth
* Confidence intervals

**Growth rate estimation**

* Phylofit
* Do this on diagnostic samples only (i.e. not the relapse samples)

**Timing somatic mutations relative to copy number aberrations**

* Used PURPLE output (CNV info) and SMuRF-filtered vcf files
* MutationalTimeR

**Targeted deep sequencing (skip this for initial version and keep it for revisions?)**

* Design of primers based on known shared mutations
* iSeq deep sequencing

**Data availability**

**Code availability**

**Acknowledgements**

**Athor contribution**

**References**

1. Mitchell, T. J. *et al.* Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell* **173**, 611-623.e17 (2018).

2. Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).

3. Roschewski, M., Staudt, L. M. & Wilson, W. H. Burkitt’s Lymphoma. *N. Engl. J. Med.* **387**, 1111–1122 (2022).

4. Gastwirt, J. P. & Roschewski, M. Management of adults with Burkitt lymphoma. *Clin. Adv. Hematol. Oncol. : HO* **16**, 812–822 (2018).

5. Minard-Colin, V. *et al.* Rituximab for High-Risk, Mature B-Cell Non-Hodgkin’s Lymphoma in Children. *N. Engl. J. Med.* **382**, 2207–2219 (2020).

6. Dozzo, M. *et al.* Burkitt lymphoma in adolescents and young adults: management challenges. *Adolesc. Heal., Med. Ther.* **8**, 11–29 (2016).

7. Woessmann, W. *et al.* Progressive or relapsed Burkitt lymphoma or leukemia in children and adolescents after BFM-type first-line therapy. *Blood* **135**, 1124–1132 (2020).

8. DALLA‐FAVERA, R. *et al.* Mechanism of Activation and Biological Role of the c‐myc Oncogene in B‐cell Lymphomagenesis. *Ann. N. York Acad. Sci.* **511**, 207–218 (1987).

9. Giulino-Roth, L. *et al.* Targeted genomic sequencing of pediatric Burkitt lymphoma identifies recurrent alterations in antiapoptotic and chromatin-remodeling genes. *Blood* **120**, 5181–5184 (2012).

10. Schmitz, R., Ceribelli, M., Pittaluga, S., Wright, G. & Staudt, L. M. Oncogenic Mechanisms in Burkitt Lymphoma. *Cold Spring Harb. Perspect. Med.* **4**, a014282 (2014).

11. Thomas, N. *et al.* Genetic subgroups inform on pathobiology in adult and pediatric Burkitt lymphoma. *Blood* **141**, 904–916 (2023).

12. López, C. *et al.* Burkitt lymphoma. *Nat. Rev. Dis. Prim.* **8**, 78 (2022).

13. Canonne, M., George, F. & Graux, C. Ex vivo culture of malignant primary B cells. *Front. Hematol.* **1**, 1004717 (2022).

14. Gonzalez-Pena, V. *et al.* Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc. Natl. Acad. Sci.* **118**, e2024176118 (2021).

15. Derks, L. L. M. *et al.* Protocol for genome-wide analysis of somatic variants at single-cell resolution using primary template-directed DNA amplification. *STAR Protoc.* **6**, 103499 (2025).

16. Middelkamp, S. *et al.* Comprehensive single-cell genome analysis at nucleotide resolution using the PTA Analysis Toolbox. *Cell Genom.* **3**, 100389 (2023).

17. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **49**, 764–766 (2013).

18. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

19. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nature Methods* **12**, 380–381 (2015).

20. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

21. Kozlov, A., Alves, J. M., Stamatakis, A. & Posada, D. CellPhy: accurate and fast probabilistic inference of single-cell phylogenies from scDNA-seq data. *Genome Biol.* **23**, 37 (2022).

22. Panea, R. I. *et al.* The whole-genome landscape of Burkitt lymphoma subtypes. *Blood* **134**, 1598–1607 (2019).